



Économie et institutions

16 | 2011

Neuroéconomie et interactions sociales

The biological basis of intertemporal choice and the design of debiasing institutions

Sacha Bourgeois-Gironde and Itzhak Aharon



Electronic version

URL: <http://journals.openedition.org/ei/106>

DOI: 10.4000/ei.106

ISSN: 2553-1891

Publisher

Association Économie et Institutions

Printed version

Date of publication: 30 March 2011

Number of pages: 121-140

ISSN: 1775-2329

Electronic reference

Sacha Bourgeois-Gironde and Itzhak Aharon, « The biological basis of intertemporal choice and the design of debiasing institutions », *Économie et institutions* [Online], 16 | 2011, Online since 31 January 2013, connection on 20 April 2019. URL : <http://journals.openedition.org/ei/106> ; DOI : 10.4000/ei.106

Revue Économie et institutions

The biological basis of intertemporal choice and the design of debiasing institutions

Sacha Bourgeois-Gironde* et Itzhak Aharon¹**

1. Introduction

We will argue that psychology and neuroscience not only provide a better understanding of why choices may be intertemporally inconsistent, but also give insights in the way institutional design and policies-making may reduce the negative aggregate impact of this widespread behavioral anomaly. We will first describe this anomaly and whether it can, along with other documented behavioral anomalies, have an impact at the aggregate level. It is not necessarily the case that all individual behavioral or cognitive biases add up to a collective effect. In particular they may be eliminated through markets interactions or mechanisms. Our question is whether it is putatively the case with hyperbolic discounting, i.e. the typical functional form that describes our discounting future utilities in making intertemporal choices, leading to inconsistent preference reversals. More generally we can consider that institutions, which typically solve inter-individual difficulties, may also help solve internal conflicts, as the ones which stem from lack of self control, hyperbolic discounting and the ensuing substitution of short-term detrimental behavior to long-run beneficial plans. In the second section we will link up with normative discussions about so-called “soft paternalism” (Camerer 2003; Thaler and Sunstein 2003), which can be defined as the design of institutional contexts inducing better choices by individuals without infringing on their autonomy. We will defend a version of soft paternalism. Inconsistent intertemporal preferences raise a thorny issue for soft paternalism, if we consider that those behavioral patterns may have evolved in humans (as well as in other animals) because they were advantageous on the whole. Is there, in consequence, a possible optimal institutional design to eliminate hyperbolic discounting, given that, on the one hand, individuals may fail to acknowledge the sub-optimality of their behavior, and, on the other hand, that the paternalistic interventionist may ignore local circumstances justifying a preference for the present over the future. In the third section we wonder whether social neuroscience studies of intertemporal choice may clarify those issues. We indicate the

¹ * Institut Jean-Nicod (ENS-EHESS), Paris, France.

** Interdisciplinary Centre, Herzliya, Israel.

neural mechanisms that underlie intertemporal choice. Humans are unique in the sense that their brains implement two competing systems: one yielding to impulses in the very short run, and one that discounts the future in a regular way, lending credibility to a quasi-hyperbolic model as the one proposed by Laibson (1997). The problem, though, is with conflicts between present and future maximization of self-utility: the dualistic neural account, in itself, does not provide a solution for preferring one perspective (present or future) over the other. It is left to the paternalist to define what temporal perspectives vis-à-vis individual behavior is the normatively encompassing one. By doing so, he may be led to deserve the “softness” qualification to a lesser extent.

2. Within-individual biases and institutional design

Intertemporal choices are decisions with consequences that occur or spread at different moments of time. Those decisions of course cover all aspects of life: health, food, occupation, leisure, etc. The idea of intertemporal choices is that they imply trade-offs between present gratifications and later ones. Economists have modeled those life situations initially and prevalently using the discounted utility model proposed by Samuelson (1937). In that model, individuals evaluate the benefits and losses that may result from a decision in an exponential way: the utility of consequences is continuously discounted at a constant rate. This model has been used as a tool for public policy. The discount rate used to analyze a given trade-off will guide the policy. Psychological research (e.g. Ainslie 1975) has challenged this initial model by showing that humans tended to discount time in a hyperbolic way rather than an exponential one. Exponential discounting cannot account for the frequently observed inconsistency of preferences across time nor make sense of the frailty of human plans, projects, and resolutions.

There has been opposed contentions among behavioral and experimental economists as to know whether documented biases and anomalies, when observed at the individual level, had any significance at the aggregate level, which is the one macroeconomists and policy-makers are interested in. This issue was lying at the core of a debate between Kahneman and Smith in the early 1990s. To schematize its terms the opposition was between experimental economists insisting that monetary incentives and immersion on markets tend to eliminate errors, anomalies or biases (Harrison 1992; Smith and Walker 1993) and psychologists who downplay the evidence that such incentives suffice to extirpate biases (Tversky and Kahneman 1992). The latter concede that remuneration eliminates some errors but the debate is precisely about why some errors can possibly resist market situations. Smith suggests a general answer. If

decision costs are high but rewards are potentially low, it is in the subject's interest not to be too accurate in his decision. Now, decreasing opportunity costs and increasing rewards, according to Smith's experimental evidence, tend to eliminate errors. In an influential recent paper Edward Glaeser (2004) synthesizes the point by contending that if people have the ability to correct their errors, they will spend more to correct these errors when the cost of these errors is higher. Reciprocally this means that errors would be particularly prevalent in situations where the cost of the poor assessment of one's cognitive status is low. The elimination of hyperbolic discounting anomalies along Smith and Glaeser's proposal would amount to make normative decisions less costly. But in what monetary sense is overcoming choices leading to bad future consequences costly? It is hard to define a decision opportunity cost for many biases and anomalies that have been documented. Moreover this policy would be ineffective with naïve hyperbolic discounters who may fail to understand that they will be subject to biases in the future and may think that they won't have to incur the cost of wrong choices².

It is hard to give a principled answer to the question as to which anomalies have presumably an impact at an aggregate level. As we recalled above, in the case of intertemporal choice anomalies, aggregated individual anomalous patterns of behavior presumably entail collective inefficiency, in connection with issues such as savings (Thaler and Benartzi 2004) or education and development (Mullainathan 2005). Institutions reduce externalities, help solve asymmetry of information and coordination problems, and in general have been deemed by economists to implement social norms when, in a more parsimonious economic ontology, social norms have not simply been identified with institutions. To some extent institutions are coextensive with aspects and patterns of interpersonal behavior. "Market" for instance is an institution which can receive as an extensional definition the sum of trading activities run by individuals in a given geographical and historical context. More specific institutions grafted on free markets may diversely determine trading activities, other activities than trade may be regulated as an effect of institutional manipulation, and emerging norms may simply be denominated in terms of that intentional institutional device.

What remains constant over this alternative ontology of institutions (the intended design *vs.* or *as* the patterns of behavior correlative to that effective design), is that relations between individuals, rather than phenomena taking place within individuals, appear to constitute the basic unit of institutional epistemology and practice. As Mullainathan aptly put it: "this focus on resolving

² The authors are indebted to Milo Bianchi for this particular argument.

problems *between* people, rather than *within* individuals, is natural to economists” (Mullainathan 2005). This raises the same issue as before: psychological errors have in principle little importance as long as they do not transpose into a collective problem and as long as they can be solved through repetitive interactions in a market or through the application of more specific institutional mechanisms. This position shapes a prejudiced view of what ingredients of human psychology are worth being taken into account by economists as targets of deliberate public policies. The prejudice may come in several guises. Most often economists have been accused of entertaining an unrealistic view of individuals’ cognitive abilities, selfish motivations, and unflinching desires and will. But it does not suffice to invert the terms of this prejudice, as it has also too often been proposed that by postulating a bounded cognition and willpower and ingrained altruistic motives we automatically get a more faithful view of the psychology underlying economic behavior.

The picture may be more mitigated. Imagine a mother whose perfectly lucid goals consist in favoring her child’s education through schooling in a developing country. She has been saving money for many years in order to fulfill that goal. She has reasonable, not too low, nor too high, expectations of how schooling will improve her child’s living conditions, and maybe hers in the future. She is then altruistic within the limits of an extended selfishness. Her willpower also is maximally exerted. She has resisted so far a gamut of temptations and actual needs to spend the spared money for pressing urges. Why postulate bounded cognition and willpower in her case? This is certainly not the most salient explanatory factor in the context at hand that will make sense of that mother’s eventual failure to grant her child a durable education. Environmental factors that led the mother to spend her savings for even more pressing urges (like paying her husband expansive medical care) might have made her divest from her primary goal in spite of mental lucidity and relative firmness. These are the types of cases envisioned in Mullainathan’s prospective article on how development economics may benefit from psychological and behavioral approaches (2005). The fact is simply that in some cases agents stray away from their self-declared goals not by lack of the adequate cognitive abilities but lack of possibilities to exert them fully. The mothers had to change her initial plans due to an adverse shock and by doing so she does not infringe any principle of rationality. She could have stuck to her initial plans had the budget for the education of her son be engaged in a different way.

But why not consider temptations, sudden lack of willpower, hyperbolic discounting, as sorts of adverse internal shocks? Hyperbolic discounters change their plans while no new information has arrived or a shock occurred. Even if, contrary to Mullanaithan’s

“mother”, those individuals have exemplifying typical bounded rationality or willpower, why not treat their weakness and errors on a par with within-individuals conflicts or failures which are not due to any cognitive boundedness and can, presumably, be targeted by appropriate institutional settings. The hypothesis here is that similar contextual changes can address internal conflicts due to a lucid and rational or to an impulsive and irrational move from the individual. Mullainathan (2005) makes institutional suggestions that seem to us to apply to both types of cases: spread the pressure of a costly decision (investment) over time so that it becomes harder to forego one’s original plans once the, usually punctual, decision has been made; multiply the occasion of rendering a final outcome of a long-term beneficial solution salient, as saliency of those outcomes may motivate individuals to pursue them; distill benefits of decisions in advance so that those benefits stop to be fully unattained objectives. But there may also be systematic bounds on cognition, will and motives, whose effects may be pervasive in a community and the more specific question is, with respect to those actual bounds, to know whether there are some adequate settings to limit their suboptimal impact. In Smithian terms, the type of solutions just suggested, should eliminate self-control problems as it decreases the opportunity cost of implementing a long-term beneficial solution and makes the reward more salient and more immediate, if not higher. If so, there is no further practical question to raise about systematic biases than about circumstantial hurdles to realize one’s will as both types of errors are treated in the same institutional way. But we may doubt that self-control problems can be as conveniently disposed of, for the main reason that they may have a biological reality (and evolutionary justification) which makes them less tractable than more superficial errors.

Answers to this question may contribute a better understanding of the purpose and potential efficiency of proactive institutional designing in addressing within-individuals, rather than between-people, issues. Our more specific purpose in the following is to assess how neural data, especially those providing a neurobiological understanding of intertemporal choice behavioral anomalies, may help provide those answers. But we must first recall recent twists in the debate about paternalism, the position which consists in favoring people’s self-interest and capacities in spite of their apparent lack of will or ability to act accordingly. There are several conceptions about which way of practicing benevolence is the least intrusive and the most efficient. Intertemporal choice seems the staple example of controversies about kinds of paternalisms as it blatantly encapsulates a potential conflict between short-term and long terms interests. Moreover, anomalies affecting intertemporal choice show that people have inconsistent preferences or, at least,

mutually inconsistent compartmentalized temporal windows with regard to the expression of their preferences, which makes those anomalies a privileged hard case for the institutional dealing with within-individual problems.

3. Variations of soft paternalism and their difficulties

Soft paternalism suggests that institutional designs are possible which can help individuals make the most rational choices for themselves without infringing on their autonomy (Camerer 2003; Thaler and Sunstein 2003). Such institutional designs would be benign for rational people and therapeutic for those who succumb to allegedly nefarious anomalies such as hyperbolic discounting, namely the propensity to make impulsive choices and entertain an imperfect or insufficiently compelling representation of their future interests.

Shaping public policies may legitimately follow the twin goals of preserving individual freedom and maximizing collective efficiency. But it should not impose a particular normative vision of their future on individuals and only reasonably enhance their capability to stick to their resolutions. The difficulty is the following: the preference for the present may well be a behavioral law, anchored in evolutionary motives, and, as such, optimal on the whole, which it may be vain, and irrational at a meta-level, to thwart. Only the most pathological cases leading to severe instances of social dysfunction should be dealt with, in a way, though, that may lead us astray from soft guises of paternalism.

In order for paternalism to be a legitimate policy it thus has to encompass a normative standpoint, even though it does not seek to impose it on individuals. A paradox may loom at this juncture and threaten the coherence of paternalism. In absence of such a normative standpoint, paternalist policies lose their gist and legitimacy. Now, let's suppose that there is enough theoretical and empirical ground for paternalists to define normative behavioral standards; two attitudes may ensue. Strong soft paternalism, so to say, may see to it that individuals adhere to these normative standards, so that they would gain cognitive autonomy. Weak soft paternalism may more modestly look for punctual error corrections without intending deeper mental reform on the part of individuals. Where paradox looms is that the most ambitious form of paternalism may rely, in order to achieve its goal of fostering cognitive autonomy, on not so soft settings, while weaker approaches may keep soft all the way long but become permanent. In both cases a risk of slippery slope toward institutional hardening or encrusting arises, one which had not been clearly identified in similar critiques that have been addressed to soft paternalism. But even more fundamentally, it is not

so clear that one can easily agree on a normative standpoint in the face of intertemporal choice anomalous behavior.

A reasonable assumption would be to adopt a normative standpoint with which the agent would himself tend to agree. But this agreement often comes too late, when the harm is done, and it may then essentially consist of regrets. As Paserman (2004) states: “with hyperbolic preferences, a long unemployment spell could be due to 'bad' choice; with the agent agreeing that his own choices are undesirable from a long-run perspective. Therefore, an intervention that brings the agent to choose actions that are in line with his long-run preferences may actually be welfare improving”. It may be far from easy to make the agents anticipate long-term consequences of his actions in absence of actually felt regrets of having performed them. But if their performance is necessary for feeling those regrets, the point of a paternalistic guidance becomes empty and agents would be on the whole better off by being left alone.

The psychological impact of paternalistic intervention all depends, in fact, on the import of viscerally felt consequences by contrast with purely anticipated ones. If one can agree on a long-term goal on purely rational basis, stick to that basis and overcome temptations when they present themselves, it is fine enough with the social cost of providing such information. But it is not so clear that an agent, even in a cold state of rational deliberation based on explicit information about future alternatives, would choose a life path deliberately deprived of irrationalities and “hot states”. This is a question that holders of paternalism of all brands should address: how can we know in advance that we won't value less a life whose apparent higher quality has been externally determined than a less qualitative self-determined one? Of course soft paternalism does not bind individuals but modifies the contexts in which choices are made so that long-term beneficial ones normally self-impose. But the deprivation of bad choices may also have a cost – for example that of impeding individuals of having overcome them by themselves – and it would be a simple experiment to run, in the vein of those prized by behavioral economists who inspire new paternalism, to measure hedonistic states associated with repeated choices in two contrasted situations: one in which a “good” choice is made among other good choices, and one in which it is made among choices of a more mixed nature, some of them presenting short-term satisfaction and long-term regrets.

Another example, by Gruber and Mullainathan (2002), suggest that smokers with hyperbolic discount rate would like to smoke less, but cannot. So taxes could help people deal with their self-control problem having a positive impact on the present discounted value of happiness. The condition of normative agreement is in principle fulfilled in that case. In order for policy makers to

discuss the time preferences 'problem' at least two topics should be explored: are individuals aware of their preferences inconsistency? And if so, would one agree that one's own choices are undesirable from a long-run perspective? Second, should policy makers give priority to the long-run time preferences, at the expenses of her immediate tastes, modifying the present individual's choices to make them time consistent?

One should indeed reflect on the potential value of short horizon choices, and how preference for the present may have evolved under certain environmental constraints; the question being, as we will make more precise below, to know to what extent these behavioral patterns may still continue to provide potentially apt responses to features of current social economic environments.

Another fundamental, and theoretically deeper, difficulty for paternalism may stem from models of poor self-control and addiction that account for the rationality of those behaviors. Becker and Murphy (1988) have presented the standard rational addiction model. This model excludes within-individual irrationality (and consequently any form of paternalistic policy that makes its aim to target such form of irrationality) and predicts that consumption will depend on prices of addictive goods and governmental regulation of those goods should only depend on interpersonal externalities. Gruber and Köszegi (2001) have tested these predictions, and while they are empirically confirmed, these authors also note that another formulation of the Becker-Murphy model would accommodate the facts as well. Namely, if one incorporates – in conformity with an intuitive and plausible account of deficient self-control and addiction phenomena – inconsistent preferences in the model, this enriched model continues to yield the standard predictions of prices/taxes sensitivity of addictive goods consumption, and, although less parsimonious, it possesses a finer intuitive flavor. Gruber and Köszegi's position is then, in an expected way, to draw normative implications which support paternalism, i.e. the designing of institutional devices that would specifically target "internalities", rather than externalities, imposed by addictive agents on themselves. They thus recommend a one dollar surplus tax over the one calculated through Becker-Murphy estimates – the social price so to say of aggregated internal inconsistencies.

Once again, predictive power may not be sufficient to grant behavioral models supporting paternalism firm enough theoretical ground to favor long-run preferences as a normative anchor. It is clear that blatantly inconsistent preferences create a difficulty for rational choice theory, but incorporating them (behavioral economists' fetish verb) says in itself nothing about which of mutually inconsistent preferences provides the correct normative standpoint. This choice is in fact a principled one, one with which it

may be generally the case agents agree, but not fully motivated by internal reasons. To be more precise, there are two possible levels of indeterminacy that make more complicated than usually thought the selection of a normative standpoint when facing time inconsistent preferences. The first indeterminacy relates to the numbering of discount rates in accounting for intertemporal choice. How many values of discount can there be? There may be so many, or none in particular, that one can doubt the existence of a tractable functional account of intertemporal choice, shedding doubts on the historical process of finding a seemingly adequate functional account of time-inconsistency. In second, even if functional legibility can be attained and his time-preferences well identified by the agent, will he necessarily acknowledge, even in retrospect, the normative compelling nature of the most regular (typically the one reflecting more patience) functional form? Let's address in turn those two difficulties.

Behavioral economists have tended to encompass intertemporal choice and its anomalies within a unifying functional approach. The idea that the value of a good depends on the timing of its consumption was already present in the economic thought of the 18th century but discussed in more details by who is in consequence considered the father of intertemporal choice modeling: John Rae (1834, 1905). According to Rae, someone's time preferences are explained by his "effective desire of accumulation". In 1884, Eugen von Böhm-Bawerk (1890) claimed that this systematic tendency to underestimate future pleasures may be due to humans lacking the capacity to make a complete picture of their future wants, especially when it comes to remotely distant ones. Fisher (1930) – who announced the basic economic relations in intertemporal choice – continued this approach suggesting that every person has his own rate of 'impatience', one that depends upon objective factors (size and risk of future income) and subjective factors (foresight, strength of will, habit, uncertainty, selfishness, influence of fashion). Paul Samuelson (Samuelson 1937) was the first economist who suggested the discounted-utility (DU) model, a mathematical function describing time preferences in general assuming that "the individual behaves so as to maximize the sum of all future utilities".

Many economic decisions involve outcomes that occur at different points in time. To model such decisions, discounted utility models are typically used. These models combine a utility function that reflects attitudes towards outcomes and a discount function that captures the effect of the passage of time. The most widely used discounted utility model in economics is constant discounting in which the discount function is determined by a constant rate of discount. However, much empirical research has underlined "various inadequacies of the DU model as a descriptive model of behavior" a

phenomenon referred to as decreasing impatience (Frederick et al., 2002, Read 2004). One of the major anomalies is that discount rates appear not to be invariant over different horizons. Indeed, as already noted by Strotz (1956), agents appear to discount the future relative to the present more rapidly than they discount between different dates in the future. According to this hypothesis, “people are impatient at present, but claim to be patient in the future.” (Nir, 2004).

These findings have led to the development of alternative discounted utility models, commonly referred to as hyperbolic discounting. The hyperbolic discounting models are consistent with decreasing impatience and have become quickly popular in economics. Today many applications are based on hyperbolic discounting, in particular on quasi-hyperbolic discounting: a model that was first proposed by Phelps and Pollak (1968) and made popular by Laibson (1997). In his work Laibson (1997) suggested a discrete time discount “quasi-hyperbolic” function which captures the key property of hyperbolic discounting in a more tractable functional form: preferences at time t are inconsistent with preferences at time, implying a gap between one’s long-run goals and short-run behavior. This gap may elicit some type of behavior that should be invested by policy makers (for example, retirement issues and procrastination).

But doubts have been raised on the existence of a regular functional form. Rubinstein (2003) has forcefully made the point that the same experimental evidence that refutes constant utility discounting undermines hyperbolic discounting in the same manner. Rubinstein emphasizes a basic indeterminacy in the choice of a functional form to account for our intertemporal choice behavior when we inductively try to build this functional form on empirical evidence, be it behavioral or psychological. For a behavioral economist to “incorporate” psychological underpinnings of manifest discounting behavior in his theory and make them fit with a given functional form it is necessary to understand how individuals mentally apprehend trade-offs between immediate and future larger rewards. There is no reason to find functional regularity in the way those mental trade-offs proceed. What Rubinstein sees there is the use of a similarity heuristics quite comparable to the one used in balancing rewards and their probability, and generating, in some cases anomalous behavior (Tversky 1977). Namely, when the comparison to two rewards lying at different times involve rewards that do not differ drastically and times that are far away from the other, the mental eye will focus on times rather than not very distinct rewards. The reverse is true too. Overall, it does not draw up a functional form, let alone a quasi-hyperbolic one. Rubinstein then opposes a procedural and a functional version of utility discounting.

Is the quasi-hyperbolic functional form an over-simplification of the phenomenon, making it too easy to choose, from an external viewpoint, between two obvious opposed discount rates? As we insisted, the paternalist hopes a legitimating coincidence between this external viewpoint and the agent's debiased appraisal of his inconsistent preferences. The problem is that the agent may be lucid, invert his long term preferences in favor of a seemingly impatient and poorly controlled choice, and be rational in doing so. The mother in our initial example had those characteristics. Does what has been planned in the long term necessarily prevail, in terms of compelling normativity, over sudden decisions that contradict those plans? The answer is a clear "no!" The reason is simple enough and normatively compelling: there is an asymmetry of information, with respect to seemingly local impatient decisions lying in contradiction with stably defined plans, between the observer and the agent. The paternalist observer tries to find a point of agreement between the agent and himself about the normativity of long temporal horizons and he may rally the agent to his prospects. However, he may systematically lack insights about local and sudden inversions of the plans which may however depend on even more compelling reasons than those plans. As Daniel Read convincingly puts it: "The information available to the acting-agent about the local consequences of a specific choice will often be better than the information available to the pre-agent [the planning agent]. When a dieter changes his mind and has tiramisu after promising not to, it might be because he is weak-willed, or it might be because he has only now realized how appealing the tiramisu is". (Read 2006). Or as Mario Rizzo and Glen Whitman sum up: "When the short-run benefits are closer at hand, [the agent] may have the benefit of superior local information" (Rizzo and Whitman 2009). It is clear that there could not be simple ways to reach coincidence of viewpoints, between agent's and observer's, about correct choices given this fundamentally asymmetric access to local normative reasons.

It is then of the utmost importance that policy-makers increase their awareness of the biological bases of behavioral patterns that they intend to reform; not that local reasons are biological ones and long term plans less dependent on a biological basis. The situation is more complex as short-term and long-term reasoning may conflict on an evolutionary basis. A first lesson that could be learnt from a biological understanding of the phenomenon is that good reasons may be spread over a variety of subjective temporal horizons. Even if one simplifies the account of inconsistent time-preferences in terms of a dual functional form, there is no principled reason, from a biological standpoint, to grant more normative power to one sector of the quasi-hyperbolic curve than to the other.

4. Biological bases of intertemporal choice

Is there an evolutionary justification to our intertemporal choice behavior and have neuroeconomics studies of utility discounting so far contributed to define those hypothetical evolutionary motives? Answers to these questions may help solve some of the indeterminacies of soft paternalism we expounded in the previous section. Although (some) animals display far-sighted behaviors (e.g., storing nuts for winter), these are typically preprogrammed and distinct from the type of spontaneous self-control observed in humans (e.g., deciding to go on a diet) Ranger et al. (2008). The almost uniquely human capacity to take the delayed consequences of our behavior into account appears to be directly attributable to the prefrontal cortex, the part of the brain that was the most recent to expand in the evolutionary process that produced humans (Manuck et al., 2003), and that is also the latest part of the brain to develop with age. Patients with damage to prefrontal regions tend to behave myopically, placing little weight on the delayed consequences of their behavior (Damasio, et al. 1994). On the other hand, having developed a large prefrontal cortex did not eradicate short-sighted behavior supported by other more ancient parts of the brain. In human brains the two possibilities coexist, and may generate internal conflicts. We may wonder that this is the case, and more precisely surmise that short-sighted behavior and its neural persistence may be due to certain evolutionary advantages.

Along this line Bavli (2009) argues that taking immediate reward may have provided an important advantage to our foraging ancestors, as it does for other foragers, suggesting that we have evolved in the first instance to act impulsively, and only more recently have begun to incorporate abstract reasoning into the equation. He concludes that nevertheless, evolutionary forces may have induced such a short-sighted mechanism precisely because snatching up immediate reward, even at the cost of sacrificing larger, future reward, is rule-rational (Aumann 2007) - it is a strategy that, in the aggregate, maximizes fitness.

A growing body of evidence suggests that a dorsocaudal region of the anterior cingulate cortex (ACC) responds to conflicts in processing stimuli of any kind (Carter et al., 1998; Botvinick et al., 2004; Yeung et al., 2004). This is consistent with findings from McClure et al. (2004) study in which they observed activity in a similar area of the ACC that was greater for decisions involving choices between immediate and delayed rewards than for choices between only delayed rewards. Such findings have been taken as evidence for a conflict-monitoring function of ACC, which serves to detect conditions requiring the recruitment of cognitive control

mechanisms subserved by prefrontal cortex and associated structures (Botvinick et al., 2001; Kerns et al., 2004). This conflict between short-term and long terms interests – controlled by the ACC – can also support Bavli hypothesis for the evolutionary process of intertemporal choices supported by the rule-rational approach to those choices (Aumann 2007).

Yielding to immediate small rewards may be evolutionarily advantageous because once a small reward is consumed, it gets out of sight and temptation and the subject can pursue its longer-term goals. If gains are easy to grab, with very low opportunity costs, their immediate consumption may enhance the pursuit of life strategies by smothering tingling appetites. Our foraging ancestors may have developed this sense of taking advantages of small rewards as they presented themselves in their environments. Neural mechanisms dedicated to the valuation of those immediate rewards may have developed in order to deal properly with scarce and random resources. In our contemporary economic environments, this neural system may still prove itself useful. The pursuit of predefined long-term goals and deflection from smaller immediate rewards may amount to honoring sunk costs or abstract principles in a counterproductive way.

Modern economic environment are labile and complex and the propensity to accept small rewards may be optimal in the face of the opportunity costs of more sophisticated strategies. It is also possible that the incorporation of long-term plans and self-projections in the far future into present decisions is more evolutionary recent than the tendency to accept immediate gratifications. From that evolutionary perspective, the preference of small immediate rewards over larger future ones is not the sign of our irrationality, but may rather reflect the conflict between two evolved rational rules: the incremental pursuit of long-term goals and the maximization of low cost immediate rewards. The conflict is dynamically solved if one considers, on the one hand, that aggregate immediate gains may add up to maximizing long-term fitness and, on the other hand, that predefined long-term goals are endogenously modified by actually made choices.

Like humans, nonhuman animals can be run in experimental paradigms in which they choose between smaller earlier rewards and larger later rewards (although animals need to learn about the rewards through multiple trials, whereas humans can simply be informed of the contingencies). Monterosso and Ainslie (1999) note that “people and less cognitively sophisticated animals do not differ in the hyperbolic form of their discount curves.” Some researchers (e.g. Herrnstein 1997; Rachlin 2000) hold the view that hyperbolic time discounting is effectively 'hardwired' into our evolutionary apparatus. However, there is considerable evidence that the time

discounting of humans and other animals relies on qualitatively different mechanisms (e.g., Loewenstein 1996, Shefrin and Thaler 1988). In support of these evidence Rangel et al. (2008) argue that while both humans and animals discount the future at dramatically different rates, both humans and animals display a common pattern of time discounting commonly referred to as 'hyperbolic time discounting'. However, they believe that while such findings do not rule out the possibility that humans and animals discount the future similarly, the quantitative discontinuity is indicative of a qualitative discontinuity. There is, in fact, considerable evidence that the time discounting of humans and other animals relies on qualitatively different mechanisms. Specifically, human time discounting reflects the operation of two fundamentally different systems, one that heavily values the present and cares little about the future (which we share with other animals), and another that discounts outcomes more consistently across time (which is uniquely human) (e.g., Loewenstein 1996, Shefrin and Thaler 1988).

Many different neurobiological microfoundations have been proposed to explain the preference patterns captured by the hyperbolic and quasi-hyperbolic discount functions. The most prominent examples include temptation models and dual-brain neuroeconomic models (Bernheim and Rangel, 2004; Gul and Pesendorfer, 2001; McClure et al., 2004; Thaler and Shefrin, 1981). However, both the properties and mechanisms of time preferences remain in dispute.

For example, using fMRI McClure et al. (2004) examined the brain activity of participants while they made a series of intertemporal choices between small proximal rewards ($\$R$ available at delay d) and larger delayed rewards ($\$R'$ available at delay d'), where $\$R < \R' and $d < d'$. Rewards ranged from \$5 to \$40 Amazon.com gift certificates, and the delay ranged from the day of the experiment to six weeks later. McClure et al. (2004) found that time discounting is associated with the engagement of two neural systems: limbic and paralimbic cortical structures, are preferentially recruited for choices involving immediately available rewards; and fronto-parietal regions, which support higher cognitive functions, are recruited for all intertemporal choices. Moreover, the authors find that when choices involved an opportunity for immediate reward, thus engaging both systems, greater activity in fronto-parietal regions than in limbic regions is associated with choosing larger delayed rewards. A subsequent fMRI study that replaced gift certificates with primary rewards (juice and water) that could be delivered instantly in the scanner replicated this pattern (McClure et al. 2007). Yet another study by a different set of authors (Hariri et al. 2006) found a similar pattern in a between- rather than within-subject study. However, Kable and Glimcher (2007) found no

evidence of separable neural agents that could account for the multiple-selves that are used to explain hyperbolic-like discounting behavior. Their finding argues strongly against the hypothesis that multiple-selves, with different discount functions, are instantiated as discrete neural systems at the proximal algorithmic level contradicts the McClure et al., (2004) study described above that appeared to support the dual-self β - δ model of David Laibson (1997), at a neurobiological level.

Social neuroscience should investigate the phenomenon of intertemporal choice in its deeper complexity. A variety of violations can actually be involved. Hyperbolic discounting may follow from distinct parameters whose labels are for instance magnitude effect, gain loss asymmetry, improving sequences phenomenon, common difference effect, etc. Those particular sources of the violations of rationality have not received a systematic treatment from a neurobiological point of view. We suggest that more than two neural systems are involved in hyperbolic discounting depending on the specific parameters that happen to contribute to discounting. This suggestion is along the same lines as Rubinstein (2003)'s procedural view of this behavioral anomaly. Once we more finely understand the neural mechanisms and the mental procedures involved in intertemporal choice institutional settings can be designed accordingly.

A second suggestion relates to the way we should conceive of the connection between the study of neural bases of intertemporal choice and soft paternalist policies. What do "we" want to change in individuals through pinpointed institutions: their long-term goals, their propensity to succumb to temptations, their cognitive abilities, or their image of themselves and their possibility to maintain it across time? Let's consider the latter possibility. Utility discounting has been conceptually described as an "unfair" treatment of different temporal stages of the self, one which could be considered as irrational and to some extent unethical (Ramsey 1928; Rawls 1971; Parfit 1984). Or, again, present utility discounting can be conceived of as a partial lack of continuity between the present and the future self. The future self is not treated on a par with the present one, maybe because that self *is not* the present one and intertemporal choice is better accounted for in terms of socially more or less distant entities entering into asymmetrical transactions (only the present self has some power over the future one). It is like playing a dictator game with oneself. Is there a neurobiological basis of this partial lack of interest for self-continuity?

Ersner-Hershfield and his colleagues (2009) have shown that the same type of neural activation in the rostral anterior cingulate cortex that differentiates between self- vs. other-judgments is elicited in the case of judgments involving present vs. future self. Moreover,

the individual differences observed in terms of relative neural activation for present vs. future self were correlated with intertemporal choice behavior. The more accentuated the difference for activations between present vs. future self, the steeper the discount curve between those two temporal instances of the individual in an intertemporal choice task. These findings lie at the core of the empirical analysis of major financial decisions in individuals' life, such as spending for the present or saving for the future.

What neuroeconomics can teach us – or confirm – at this juncture is that the ability to project oneself into the future is correlated with the way in which we deal with intertemporal tradeoffs. The elaboration of our self-image, the anticipated continuity of our persona or psyche – whatever this entity is called – and the stability of mental representations of future situations in which we figure ourselves in apparently go against our evolutionary anchored proneness to yield to immediate stimuli. But there is no normative reason to side on one way or other of this cognitive and affective divide, since there is no principled reason either that myopic choices or apparent hyperbolic behavior be always incompatible with long term goals. In particular, if long term goals are internally present in individuals' minds and if institutional environments facilitate their achievement by reducing the cost of their maintenance, there is no reason to think that yielding to impatient choices is necessarily conducive to self-destruction. It may even be understood as a diversification strategy, i.e. the punctual foraging of various gratifications which in some cases may beneficially deter the individuals from ill-defined long-term goals.

Conclusion

How neural and biological data and the confirmation they provide that specific neural bases for hyperbolic discounting may have been evolutionarily selected impact on economics? The purpose of the biological understanding of intertemporal choice is not simply to validate or invalidate alternative models of discount utility, it is, in our view, to connect more broadly the observed behavioral patterns with evolutionary motives that may explain them, with neural systems that were not initially thought of in connection with the issue of intertemporal choice, and with policies which address the issue in its whole complexity. A correlation between discount rates and a lack of interest for one's future self, as evidenced by relevant neural observations, may shift the primary target of attempts to heal impulsivity from decreasing or increasing opportunity costs for intended normative decisions to developing the ability to project into one's future self and decrease the perceived discontinuity between

the self now and tomorrow. It is an open field for new guises of paternalism to think of which institutional setting might foster this sense of self-continuity.

References

- Ainslie, G. (1975). "Specious reward: a behavioral theory of impulsiveness". *Psychol. Bull.* 82, 463–496.
- Aumann, R. (2008). "Rule-rationality vs. act-rationality", discussion paper #497. Centre for Rationality, HUJI.
- Bavli, H. (2009). "Rule-Rationality and the evolutionary foundations of hyperbolic discounting". Discussion Paper #513, Centre for the Study of Rationality, Hebrew University Jerusalem.
- Becker, G. and Murphy, K., (1988). "A Theory of rational addiction", *Journal of Political Economy*, 96, 675-700.
- Bernheim, B.D., and Rangel, A. (2004). "Addiction and Cue-Triggered Decision Processes." *American Economic Review*, 94(5), 1558-90.
- Bohm-Bawerk, Eugen von (1890) [1884]. *Capital and Interest: A Critical History of Economical Theory*. London: Macmillan Company.
- Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003), "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism", *University of Pennsylvania Law Review*, 1151(3), 1211-1254.
- Damasio H, Grabowski T, Frank R, Galaburda AM, Damasio AR. (1994). "The return of Phineas Gage: clues about the brain from the skull of a famous patient". *Science*. 264(5162):1102-5.
- Dasgupta, (2008). "Discounting Climate Change", *Journal of Risk and Uncertainty*, 2008, 37(2-3), 141-169
- Ersner-Hershfield, H., Elliott Wimmer G., and Knutson, B. (2009). "Social Saving for the future self: Neural measures of future self-continuity predict temporal discounting", *Cognitive and Affective Neuroscience*, 4(1): 85-92.
- Frederick S., G. Loewenstein and T. O'Donoghue (2002), "Time Discounting and Time Preferences: A Critical Review", *Journal of Economic Literature*, 40 (2), 351-401.
- Glaeser, E. (2004). "Psychology and the Market," *American Economic Review*, vol. 94(2), 408-413.
- Gruber, J. and Köszegi, B., (2001). "Is Addiction 'Rational'? Theory and Evidence", *Quarterly Journal of Economics*, 116, 1261-1303.
- Gruber, J. and Mullainathan, S. (2002). "Do Cigarette Taxes Make Smokers Happier?", NBER Working Papers 8872, National Bureau of Economic Research.

- Gul, F. and Pesendorfer, W. (2001). "Temptation and Self-Control." *Econometrica*, 69(6):1403-1435.
- Harrison, G., (1992). "Theory and Misbehavior of First Price Auctions: Reply," *American Economic Review*, 82, 1426-1443.
- Herrnstein R.J. (1997). "Melioration as Behavioral Dynamics" in H. Rachlin and D. Laibson (eds.) *The Matching Law: Papers in Psychology and Economics*, pp. 74-99. Cambridge Mass:Harvard University Press.
- Kable J., and Glimcher P. (2007). "The neural correlates of subjective value during intertemporal choice". *Nat Neurosci*. 10(12):1625-33.
- Laibson, D. (1997), "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, 112 (2), pp. 443-77.
- Loewenstein, G. (1996). "Out of control: Visceral influences on behaviour". *Organizational Behavior and Human Decision Processes*, 65, 272-92.
- Manuck SB, Flory JD, Muldoon MF, and Ferrell RE. (2003). "Central nervous system serotonergic responsivity and aggressive disposition in men". *Physiological Behavior*. 77(4-5):705-9.
- McClure S, Ericson K, Laibson D, Loewenstein G, and Cohen J. (2007). "Time discounting for primary rewards. *J Neurosci*." 27(21):5796-804.
- McClure SM, Laibson DI, Loewenstein G, and Cohen JD. (2004). "Separate neural systems value immediate and delayed monetary rewards". *Science*, 306(5695):503-7.
- Monterosso, J. R., Ainslie, G. (1999). "Beyond discounting: possible experimental models of impulse control". *Psychopharmacology (Berl)*. 146, pp. 339-47.
- Mullainathan, S. (2005). "Development Economics Through the Lens of Psychology" in *Annual World Bank Conference in Development Economics 2005: Lessons of Experience*, edited by F. Bourguignon and B. Pleskovic. Oxford, UK and Washington, DC: Oxford University Press and World Bank
- Nir A. (2004), "Cognitive Procedures and Hyperbolic Discounting", Tilburg University, Center for Economic Research, Discussion Paper No. 47.
- Parfit, D. (1984), *Reasons and Persons*, Clarendon Press: Oxford.
- Paserman, M.D. (2004), "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation", *IZA Discussion Paper* No. 997.
- Phelps, E.S. and R. Pollak (1968), "On Second Best National Saving and Game-Equilibrium Growth", *Review of Economic Studies*, 35, pp. 185-199.

- Rachlin, H. (2000). *The Science of Self-Control*, Cambridge; London: Harvard University Press.
- Ramsey, P. (1928). "A Mathematical Theory of Saving", *The Economic Journal*, 38(152), 543-559.
- Rangel, A., Camerer, C., Montague, R. (2008). "A framework for studying the neurobiology of value-based decision making". *Nature Reviews Neuroscience*, 9, 545-556.
- Rawls, J. (1971). *A Theory of Justice*, Cambridge Mass. Harvard University Press.
- Read, D. (2001). "Is Time-Discounting Hyperbolic or Subadditive?". *Journal of Risk and Uncertainty*, 23(1), 5-32.
- Read, D. (2004). "Intertemporal Choice". In KOEHLER, D.; HARVEY, N. (eds.). *The Blackwell handbook of judgment and decision making*. Oxford: Blackwell.
- Read, D. (2006). "Which Side Are You On? The Ethics of Self-Command", *Journal of Economic Psychology*, 27, 681-685.
- Rizzo, M. and Whitman, G. (2003), "Little Brother is Watching You: New Paternalism on the Slippery Slopes", *Arizona Law Review*, 51, 3 685-739.
- Rubinstein, A. (2003), "Economics and Psychology? The case of hyperbolic discounting", *International Economic Review* 44 (4), pp. 1207-1216.
- Samuelson, P. (1937), "A Note on Measurement of Utility", *Review of Economic Studies*, 4, pp. 155-161.
- Shefrin, H. and Thaler, R. (1988). "The behavioral life-cycle hypothesis". *Economic Inquiry* 26 (4), 609-43.
- Smith, V. and Walker, J. (1993). "Monetary Rewards and Decision Cost in Experimental Economics," *Economic Inquiry*, 31, 245-261.
- Strotz, R. (1956), "Myopia and Inconsistency in Dynamic Utility Maximization", *Review of Economic Studies*, 23, pp. 165-180.
- Thaler, R. and Benartzi, S. (2004). "Save More Tomorrow (TM): Using Behavioral Economics to Increase Employee Saving", *Journal of Political Economy*, 112(S1), S164-S187.
- Thaler, R., and Shefrin, H. (1981). "An economic theory of self-control". *Journal of Political Economy* 89 (2), 392-406.
- Thaler, R. and Sunstein, C., (2003) "Libertarian Paternalism", *American Economic Review*, 93(2), 175-192.
- Tversky, A. (1977). "Features of similarity". *Psychological Review*, 84, 327-352.
- Tversky, A. and Kahneman, D. (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297-323.